**U.S. Army Research Institute
for the Behavioral and Social Sciences**

Research Report 1875

Development and Content Validation of Crisis Response Training Package Red Cape: Crisis Action Planning and Execution

Scott B. Shadrick and Peter S. Schaefer
U.S. Army Research Institute

Jeff Beaubien
Aptima, Inc.

August 2007

**U.S. Army Research Institute
for the Behavioral and Social Sciences**

**A Directorate of the Department of the Army
Deputy Chief of Staff, G-1**

Authorized and approved for distribution:

**PAUL A. GADE
ACTING TECHNICAL DIRECTOR**

**MICHELLE SAMS, Ph.D.
DIRECTOR**

Technical review by

COL Richard Shatto, Indiana National Guard, Joint Forces Headquarters
John S. Barnett, U.S. Army Research Institute

NOTICES

REPORT DOCUMENTATION PAGE

1. REPORT DATE (dd-mm-yy)	2. REPORT TYPE	3. DATES COVERED (from. . . to)
August 2007	Final	March 2005 to April 2007

4. TITLE AND SUBTITLE	5a. CONTRACT OR GRANT NUMBER
Development and Content Validation of Crisis Response Training Package Red Cape: Crisis Action Planning and Execution	5b. PROGRAM ELEMENT NUMBER 622785

6. AUTHOR(S)	5c. PROJECT NUMBER A790
Scott B. Shadrick, Peter S. Schaefer (USARI), and Jeff Beaubien (Aptima, Inc.)	5d. TASK NUMBER 211
	5e. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
U.S. Army Research Institute for the Aptima, Inc. Behavioral and Social Sciences 12 Gill Street Suite 1400 ATTN: DAPE-ARI-IK Woburn, MA 01801 121 Morande Street Fort Knox, KY 40121-4141	

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. MONITOR ACRONYM ARI
U.S. Army Research Institute for the Behavioral and Social Sciences 2511 Jefferson Davis Highway Arlington, VA 22202-3926	11. MONITOR REPORT NUMBER Research Report 1875

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

Subject Matter POC: Scott B. Shadrick

14. ABSTRACT *(Maximum 200 words)*:

Military and civil crisis response organizations need training which accelerates the development of expertise in effective and efficient interagency cooperation and collaboration. The development of such training necessitates the use of a cognitive task analysis (CTA) method which can draw upon distributed expertise to understand not only current but future task conditions. A newly developed CTA method known as the Flexible Method of CTA (FLEX) was applied to the domain of crisis response and resulted in the training program known as Red Cape: Crisis Action Planning and Execution. Quantitative assessments of the training content validity were elicited from seasoned crisis response personnel. Analysis of the assessments supports the use of FLEX in developing crisis response training. Proposed extensions and refinements of existing content validation procedures are also discussed.

15. SUBJECT TERMS

Crisis Response Training, Interagency Training, Cognitive Task Analysis, Distributed Expertise, Knowledge Creation

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON
16. REPORT	17. ABSTRACT	18. THIS PAGE			Ellen Kinzer Technical Publication Specialist 703-602-8047
Unclassified	Unclassified	Unclassified	Unlimited	33	

Research Report 1875

Development and Content Validation of Crisis Response Training Package Red Cape: Crisis Action Planning and Execution

Scott B. Shadrick and Peter S. Schaefer
U.S. Army Research Institute

Jeff Beaubien
Aptima, Inc.

Armored Forces Research Unit
Barbara A. Black, Chief

DEVELOPMENT AND CONTENT VALIDATION OF CRISIS RESPONSE TRAINING PACKAGE RED CAPE: CRISIS ACTION PLANNING AND EXECUTION

EXECUTIVE SUMMARY

Research Requirement:

Effective military and civil interagency collaboration is integral to successful crisis response efforts. However, existing training does not readily allow for the development of expertise in such collaboration. Further, training for crisis response efforts must address future—as well as current—task conditions. Such training must also draw upon expertise distributed across individuals from different specialties. Unfortunately, traditional cognitive task analysis (CTA) methods are inadequate for the development of such training.

A successful remedy requires the development of a new CTA method, and the application of a proven training methodology which allows for the accelerated development of expertise. The new CTA method is known as the Flexible Method of CTA, or FLEX, and is expressly designed to draw upon distributed expertise in knowledge dealing with current and future task conditions. The theme-based training methodology utilized in the Red Cape training was adapted from the successful Think Like a Commander: Captains in Command program. The goal was to design suitable training for military, civil, and government agencies involved in crisis response efforts.

Procedure:

Researchers reviewed technical documents and interviewed experts to determine probable scenarios for which contingency plans must be developed. Utilizing the FLEX approach, subject matter experts (SMEs) were then interviewed to identify examples of historically effective and ineffective crisis response behaviors and further validate initial assumptions. Once the set of scenarios was finalized, a theme-based training approach (similar in nature to that utilized in the successful *Think Like a Commander* program) was used to generate a training package known as Red Cape: Crisis Action Planning and Execution. The content validity of Red Cape was assessed by quantitative indices elicited from an independent sample of Army National Guard personnel with crisis response experience.

Findings:

Both the appropriateness of training items (through the use of the Content Validity Ratio [CVR]) and the structure of the training (through the use of Cohen's Kappa) were assessed. Analysis of the CVR training item values exhibited acceptable reliability (ICC= .62, $p < .001$). Of the 122 training items evaluated for inclusion/exclusion, only one was selected for removal. Results also indicated significant agreement between the preferred training structure as expressed by the SMEs and the training structure derived from the FLEX CTA (Cohen's $\kappa = .78$, percent agreement = .85, $p < .001$). Overall, results indicated support both for the training content and

for the use of the FLEX approach to derive training content. Further issues involving the construction and validation of such training, the selection and use of appropriate experts, and the extension and refinement of content validation procedures are discussed.

Utilization and Dissemination of Findings:

The research supports the use of theme-based training to instill higher-order cognitive behaviors. It demonstrates a low-cost, effective way to maximize the ability of diverse stakeholder groups to cooperate and coordinate in a domain in which failure is costly. The training is intended to maximize benefits from participating in large scale readiness exercises. The training has been fully implemented with the Indiana Army National Guard and their civil-military, interagency partners. The training program has also been briefed to the Commanding General and Deputy Commanding General of the U.S. Army Armor Center, and the results have been briefed to representatives of the U.S. Army Training and Doctrine Command, U.S. Army Management Staff College, and to various Army National Guard units.

DEVELOPMENT AND CONTENT VALIDATION OF CRISIS RESPONSE TRAINING PACKAGE RED CAPE: CRISIS ACTION PLANNING AND EXECUTION

CONTENTS

List of Tables

DEVELOPMENT AND CONTENT VALIDATION OF CRISIS RESPONSE TRAINING
PACKAGE RED CAPE: CRISIS ACTION PLANNING AND EXECUTION

Introduction

Crisis response operations require that relevant military and civilian agencies effectively coordinate and collaborate. However, recent events such as Hurricane Katrina have made clear that attained levels of collaboration are often sub-optimal. Effective interagency collaboration in crisis response efforts is a difficult undertaking for several reasons. First, each organization tends to have its own culture and "language," thereby making clear and concise interagency communication difficult. Second is the different, and sometimes contradictory, standing operating procedures (SOPs) employed by different organizations (Dixon, 2006). Third is the fact that the dynamic nature of crisis response situations dictates that leadership responsibilities often rotate throughout the course of a mission. For example, assume there has been an explosion in an industrial plant. In the early aftermath, the primary focus is upon dousing the flames. Thus, the lead agency might be the local fire department. However, subsequent investigation indicates that the fire was deliberately set, which means that law enforcement agencies take the lead. Finally, during the cleanup phase, the Environment Protection Agency (EPA) might take the lead. Fourth is the distributed nature of expertise: any interagency effort requires the understanding of expertise that is scattered across individuals with different specialist backgrounds. Most important, however, is the lack of opportunity to train and practice these activities.

As recent research has noted, "Events that substantively bring military decision makers in contact with other agency representatives, indigenous authorities, or commercial enterprises of relevance to operations are extremely rare" (Glenn et al., 2006, pp. 200-201). Furthermore, the infrequent events which do bring these diverse stakeholder groups together are usually large scale readiness exercises. Although beneficial, those events are not training exercises per se. Rather, such exercises emphasize coalition and organizational assessment rather than individual and team learning. Therefore, in addition to being too infrequent to train expert-like behaviors, those exercises also lack the characteristics of situations which foster true expertise (Ericsson, Krampe, & Tesch-Römer, 1993). Crisis response professionals are also expected to develop contingency plans for events judged likely to happen, not just events which have already occurred. Hence, even when crisis response professionals have developed procedures for such situations, those procedures have not yet been tested in the field.

Crisis situations are not confined to natural disasters; acts of terror are also a major concern. Military and crisis strategy experts are stressing that the approach to terrorism needs to change from inefficient global policies to more effective local ones (Fallows, 2006). Those experts caution that it is impossible to prevent every threat in every locale. However, it is possible to ensure that if a crisis event should occur, response would be timely and effective (Fallows, 2006). The timeliness and effectiveness of any such response will depend upon the appropriate training.

For many domains, training can be constructed by using traditional cognitive task analysis (CTA) methods. These methods enable researchers to extract, codify, and measure

expert performance. These methods also provide contextually rich information that can be used to develop effective training programs which accelerate the development of expertise. Unfortunately, those methods are also largely suited for domains with relatively static task demands—a trait which is hardly typical of crisis response efforts.

Developing training for crisis response operations requires methods that can address rapidly changing task demands, as well as tasks that have not yet been identified (Fallows, 2006). Task demands can change rapidly because of 'on the ground' factors, or because of technological innovations, which can affect "judgments, roles, relationships, and weightings on different goals" (Woods & Dekker, 2000, p. 5).

One way of illustrating the inadequacy of traditional CTA approaches for the domain of crisis response is to make a distinction between knowledge elicitation and knowledge creation. Knowledge elicitation involves the extraction of current knowledge about task conditions from experts (Chervinskaya & Wasserman, 2000). However, when—as is the case for crisis response—training must address future, rather than current, task conditions (Knapp et al., 2002)—two further conditions must be met (Shadrick, Lussier, & Hinkle, 2005).

First, the task analysis method must be able to extract distributed expertise—that is, expertise that is distributed across organizations and individuals from different specialty areas (Meredith & Mantel, 2004). Second, knowledge elicitation alone is not enough. It is also necessary to create *new* knowledge. Knowledge creation has been defined as a continuous process for transferring and sharing tacit and explicit knowledge with groups and individuals (Bloodgood & Salisbury, 2001). The knowledge creation process results in the development of principles, facts, concepts, procedures, and processes—products that must be extended to develop future concepts (Shadrick, Lussier, & Hinkle, 2005).

In sum, the situation is as follows. First, interagency collaboration in crisis response efforts is both vital and difficult. Second, current training for interagency collaboration is both infrequent and inadequate. Third, traditional CTA approaches are insufficient for the development of such training. What is needed is a method which provides structured activities to measure, assess, and guide the conceptual development process—structured activities which are nonetheless also flexible enough to respond to a wide range of conceptual constructions (Shadrick, Lussier, & Hinkle, 2005). An adequate solution, therefore, requires two things. First, a new CTA method must be used to generate the training content. Second, the training methodology must foster the acquisition of expertise.

In this paper, we describe a recently developed method of CTA (the Flexible Method of Cognitive Task Analysis, or FLEX) expressly designed to elicit distributed expertise and to foster knowledge creation. We then briefly discuss the adoption of a well established theme-based training approach. Specifically, we describe how we applied FLEX to elicit distributed expertise and create new knowledge for the crisis response domain, and then used the resulting information to develop training, based on a proven methodology, for selected crisis response situations. In a follow-on effort, we conducted a quantitative assessment of the content validity of the training materials. Finally, we discuss future research issues in light of our findings.

To place FLEX in context within the domain of CTA, we describe the requirements of knowledge elicitation in emerging domains and how FLEX responds to them.

Knowledge Elicitation vis-à-vis Knowledge Creation

Cognitive task analyses are systematic approaches to understanding, to describing, and to defining a job in terms of the cognitive processes needed to successfully perform particular tasks. These approaches aid in the identification of the requisite knowledge, skills, and abilities that are required for successful job performance (Schraagen, Chipman, & Shalin, 2000). Because CTAs are based upon expert knowledge, knowledge elicitation is a common characteristic of most CTA methods (DuBois, Shalin, & Borman, 1995).

Knowledge elicitation is a process in which a researcher captures knowledge, procedures, and strategies through systematic interactions with domain experts. These techniques provide the expert with an appropriate set of procedures for transferring information from the expert to the researcher (Meyer & Booker, 1991). Such formalized procedures are necessary because of the nature of expertise, which is heuristic, difficult to verbalize, and not easily captured by standard methods of observation (Wei & Salvendy, 2004).

Much research on knowledge elicitation has been conducted and numerous methods have been developed (Chervinskaya & Wasserman, 2000). Due to the complexity of most domain areas and the difficulties associated with forecasting future capabilities and requirements, a combination of several knowledge elicitation methods is often employed to collect information from knowledgeable experts. For this reason, FLEX is a synthesis of proven methods of knowledge elicitation, as well as newly developed methods for fostering knowledge creation.

Methods of knowledge creation.

Unlike current task conditions that call for eliciting knowledge from domain experts, probable future task conditions require the creation of new knowledge. However, knowledge creation can only proceed once the future job domain is well understood. That is, knowledge elicitation is a necessary but insufficient step for knowledge creation. Several systematic methods have been devised that allow researchers to methodically collate, understand, synthesize, and extend current knowledge into the future. Three of the most common methods are environmental scanning, technological forecasting, and backcasting.

Environmental scanning is a systematic method of understanding the future job environment. It involves, among other things, the analysis and utilization of current trends (Aguilar, 1967). The goals of environmental scanning include the detection of relevant trends, and the implication of those trends for the problem under investigation. Technological forecasting involves predicting the types of technologies that will be available in a given future time, divining the nature of those technologies, and providing a realistic estimate of their availabilities (Meredith & Mantel, 2004). However, in the current paper, forecasting involves not just technological changes, but any change relevant to the domain of interest, including

changes in the tactics used by an adaptive enemy, changes in military doctrine, or changes in the amount or type of resources available. The current synergistic nature of technological advancement requires tapping experts from a diverse number of relevant domains (Meredith & Mantel, 2004). Backcasting is, in some ways, an extension of environmental scanning. The process begins with the identification and description of a desirable future end state (Robinson, 1982). Then the group determines what reasonable steps must be accomplished to achieve that desired end state.

Enabling Knowledge Creation

Traditional methods of cognitive task analysis and knowledge elicitation are effective tools for understanding expert performance in current tasks whose demands are relatively stable. However, they are not designed for predicting and assessing realistic observations and concepts about future tasks e.g., the kind of eventualities for which crisis response organizations must prepare. Effective methods are needed for capturing knowledge from available information sources and using that knowledge to create new ideas and concepts and describe potential future tasks.

Drawing on the distinction between implicit and explicit knowledge, Nonaka and Takeuchi (1995) presented a model for knowledge creation. Their model presents the different ways in which knowledge can be converted into new concepts and ideas. Knowledge creation is viewed as a continuous process for transferring and sharing knowledge (both implicit and explicit) between individuals and groups. However, the models of knowledge creation described by Nonaka and Takeuchi may not be appropriate when dealing with a domain undergoing rapid change. To understand expert performance in such domains, the method must be expanded to include steps for creating unique knowledge.

The FLEX is, at its core, an interview-based problem solving approach that systematically develops and explores future concepts and the associated tasks. As noted above, this methodology is a synthesis and extension of existing approaches to knowledge elicitation and newly developed approaches to knowledge creation. Unlike existing methods, however, FLEX grounds the experts' thinking in a futuristic setting. Knowledge is then captured by employing a vignette-based scenario approach. In this approach, a potential future situation is presented to experts, who are then required to solve complex problems using anticipated resources.

The focus of FLEX upon the development and evaluation of future knowledge differentiates it from traditional CTA and knowledge elicitation methods. The FLEX is also more comprehensive than traditional CTA methods in that it not only gathers traditional task information, but also provides a methodology that can capture the conglomerate of interdependent and dynamic knowledge sources. The method allows researchers to capture existing knowledge and facilitates the creation of new knowledge and concepts.

The FLEX begins by presenting experts with a problem space. The experts are then required to generate a solution to the problem utilizing anticipated resources and domain knowledge. Participants are asked to verbalize their responses by thinking aloud, in a fashion

similar to Protocol Analysis (Ericsson & Simon, 1980; Wright & Ayton, 1987). However, in common with the Delphi technique (Dalkey, 1969), responses elicited in earlier phases are provided to experts in successive phases. Thus, subsequent participants are able to identify weaknesses, confirm strengths, and build upon the prior responses. Throughout the process, a semi-structured interview (Meyer & Booker, 1991) is used to probe experts' knowledge and gain a deeper understanding of the experts' reasoning. Before the completion of the process, responses from subsequent participants are presented to experts from earlier phases. Finally, a small group of experts is used for interactive group discussions to foster consensus building and validation.

The scenario-based approach is appropriate in this context because it is often difficult for individuals to speculate on how future capabilities and resources might be employed. Further, many experts tend to overestimate the impact of future technological advances and to ignore the difficulties. Thus, forcing the participants to solve a concrete problem helps ground their thinking so that more appropriate and realistic information can be captured. The phases by which FLEX accomplishes these goals are detailed below (from Shadrick, Lussier, & Hinkle, 2005).

Phases of FLEX

Phase 1: Domain and problem identification.

In the first phase of *FLEX*, it is necessary to define the domain area and identify the significant problems of interest. Defining the domain helps to specify the scope of the information that will be needed and assists in the identification of experts. Because a goal of the method is not only to capture new knowledge but to systematically investigate the efficacy of the information collected, the identification of a significant number of appropriate experts is required. In most cases, a diverse set of experts from a variety of highly specialized domains will be required, particularly when dealing with situations wherein the relevant expertise is widely distributed. During the problem identification phase, an environmental scan, backcast, and/or technological forecast is conducted to assist in understanding the potential for innovation and advancement, as well as changes in any other relevant factors.

Phase 2: Initial review and analysis.

During the initial review and analysis it is important to capture information related to current state-of-the-art knowledge and resources for the domain and problem space. This phase is similar to what would be expected during traditional task analysis, both procedural and cognitive. During the initial review it is necessary to interview experts, document current processes, and gather information about expert performance. Contrary to traditional task analysis, the goal is not to develop a complete list of tasks and duties in a given domain. The goal is to gain an understanding of domain expertise, task requirements, processes, and outcomes. During the review it is critical to identify areas where innovation may lead to improved performance and processes and to gain an understanding of impending innovations in the domain.

Phase 3: Refine problem space and develop initial scenario.

During this phase, initial decisions regarding the knowledge elicitation scenario are established. An initial group of domain experts is used to develop one or more scenarios with appropriate "branches." The branches represent contentious areas where multiple outcomes are possible. The scenarios are developed in an iterative fashion, allowing experts to develop a realistic situation capable of focusing the knowledge elicitation process. During this phase, it is only necessary to establish the initial conditions for the knowledge elicitation process.

Phase 4: Initial knowledge elicitation.

Based on the scenario, this phase elicits knowledge and concepts through a series of interviews with experts from relevant domains. The scenario is used as a starting point to focus the experts on a particular problem space. The elicitation process allows the experts to further refine the scenario by adding new information, challenging existing assumptions, anticipating unintended consequences, and predicting future advancements. During the elicitation, experts are asked to discover new ways to solve problems associated with the scenario given hypothesized future resources. Knowledge elicitation focuses on both individuals and small groups and combines experts from different specialty areas.

Phase 5: Data reduction and consensus building.

After a sufficient number of experts have had the opportunity to complete the knowledge elicitation phase, it is necessary to aggregate the data into common and meaningful responses. The data reduction phase allows a new group of experts to develop a consensus on the efficacy of the new knowledge and concepts captured in Phase 4. The information is then used to update the scenario to reflect the new knowledge stream or create a new branch to highlight "what if" situations. Altering the scenario allows the experts to "leave their fingerprints" on the scenario and supports an iterative process for continuous improvement. The process also allows domain experts to evaluate their organizations' SOPs for robustness in a variety of scenarios.

After revising the scenario based on the initial knowledge elicitation and data reduction phases, it is necessary to reexamine the scenario with a new set of experts and provide the revised materials to former participants. This means, at a minimum, that two independent samples of experts are required. This process provides a systematic way to evaluate the realism of the new information. Thus, Phases 4 and 5 may be repeated several times until an agreed upon solution is developed. The goal of the iterative process is to develop an expert identified and documented solution for the purpose of developing new theories, principles, tools, techniques, and procedures.

Phase 6: Knowledge representation and concept documentation.

Knowledge representation provides a mechanism for documenting and displaying information in a useable format. In this context, knowledge refers to organized concepts, theories, principles, descriptions, and mental models of descriptive, procedural, and meta-

cognitive information. The goal is to present the results of the knowledge elicitation process in a meaningful way.

The Think Like a Commander Approach to Deliberate Training

As noted earlier, the development and application of an appropriate CTA method must be accompanied by a training methodology which can accelerate the acquirement of expertise. Training designed to foster the acquisition of expertise should incorporate certain principles (Ericsson, Krampe, & Tesch-Römer, 1993; Charness, Tuffiash, & Krampe, 2005). These principles include (Lussier, Shadrick, & Prevou, 2003):

1. Knowing what expert behavior "looks like."
2. Performing the task and noting discrepancies from the expert model.
3. Repeating the task and successively approximating the expert model through conscious attention (i.e., coaching).
4. Repeating the task until behavior is habitual.
5. Once criterion performance is achieved, repetition without conscious attention.

These principles (henceforth referred to as "deliberate training") can be used to train patterns of cognition. For example, in the Think Like a Commander (TLAC) program, these principles have been successfully used to train tactical thinking behaviors in Soldiers (Lussier, Shadrick, & Prevou, 2003; Shadrick & Lussier, 2004; Shadrick, Lussier, & Fultz, 2007). Although Soldiers may possess the requisite knowledge for successful tactical thinking under time and resource constraints, such knowledge is a necessary but not sufficient condition for performance under such conditions.

The goal of the training approach was not to provide the Soldiers with the requisite background knowledge (Lussier, Shadrick, & Prevou, 2003), but rather to ingrain automatic thinking patterns characteristic of expert tacticians. Such automaticity allows for more flexible and comprehensive understanding of the battlefield, because more cognitive resources are "left over" for other activities.

The TLAC uses a theme-based training approach. In theme-based training, CTA methods are used to isolate and identify expert behaviors associated with performance in a cognitive task (Lussier, Shadrick, & Prevou, 2003). Those expert behaviors are then grouped into recurring patterns, or themes, which are broadly applicable across a range of situations. The themes are used to develop a set of scenarios (i.e., tactical situations). Within each scenario are expert indicators which must be considered before successful tactical thinking can occur. While the indicators themselves are unique to a given scenario, the indicators also belong to specific themes. For example, although the theme "Consider Effects of Terrain" is appropriate in a variety of scenarios, the expert indicators which belong to that theme vary from scenario to scenario. In other words, the expert indicators represent how a given theme is appropriately reflected in a given situation.

The scenarios are presented to Soldiers who are required, within a set time frame, to attempt to identify those indicators. Once a scenario is finished, the expert-derived set of

indicators and themes is provided to the Soldier and comparison between the Soldiers' responses and the expert responses is then possible. Upon occasion, instructors will lead classroom discussions making more explicit the decision making process of the experts which led them to that particular set of indicators. Soldiers then proceed to the next scenario. The goal of the training is to provide Soldiers with the opportunity to see how experts size up tactical situations, to apply those expert behaviors, and then through further practice and corrective feedback to automatize the expert behaviors in their own thinking.

There is experimental evidence demonstrating that the TLAC training does promote increased proficiency in such tactical exercises, even under increasingly stringent time constraints (Shadrick & Lussier, 2004). The importance of that military skill is supported by the findings of Shadrick, Lussier, and Fultz (2007) which demonstrated that TLAC performance increases as a function of officer rank and experience. Further, those with deployment experience in Operation Iraqi Freedom (OIF)/Operation Enduring Freedom (OEF) outscore the non-deployed. Therefore, the role of deliberate training in the development of thinking behaviors is supported by experimental and empirical evidence. In addition, non-deployed captains who received TLAC training performed better than both a sample of captains who had been deployed up to one year as well as a sample of lieutenant colonels (Shadrick, Lussier, & Fultz, 2007). That is a pertinent point for the current research, because it demonstrates the ability of such theme-based training to aid in the accelerated development of expertise. Similarly, scholars in other fields are beginning to apply the principles of deliberate training to complex cognition (e.g., legal studies; see Farmer & Williams, 2005).

It should also be noted that several parallels between combat situations and crisis response efforts exist. Both combat and crisis response efforts require the successful comprehension and integration of multiple, competing demands upon the cognitive system. In both cases, training is often not available which fosters the development of automatic, expert like patterns of thinking. In both cases, such training would be quite beneficial, both because the automatization of such thinking patterns would free up much needed cognitive resources and because of the high price of failure. When these parallels and the effectiveness of the TLAC approach are considered as a whole, the decision to adapt the TLAC approach to the domain of crisis response seems highly appropriate.

Implementing FLEX in Crisis Action Planning and Execution

A key strength of FLEX lies in its unique ability to capture, create, and organize the knowledge of domain experts regarding events that have yet to actually occur. This is important, because the development of effective training for homeland security, natural disasters, or any crisis response situation poses several unique problems. First, developing an effective response to such events requires the coordinated response of multiple government agencies across all levels of government (Executive Office of the President: National Science and Technology Council, 2006). For this reason, multiple groups of domain experts from diverse agencies must be involved in the CTA and training design processes (Landis, Fogli, & Goldberg, 1998). Second, many so-called experts are not true "experts" in the strictest sense of the term. Specifically, while these individuals may understand their own agencies' SOPs for responding to such events, few have actually experienced them first-hand. As a result, although they may have

the necessary factual knowledge, they have not yet developed the automatized thinking patterns that are the hallmark of a true expert. Finally, given agencies' extremely tight training budgets, there are often few opportunities to practice and receive process oriented feedback on the critical behaviors that are required for executing a combined military-civilian interagency response (Glenn et al., 2006). As a result, for training to work in this domain, it must be provided at low cost and on demand.

To address this critical requirement, FLEX was applied to the domain of crisis response situations. Following the capturing of distributed expertise, the resulting information provided the content for a prototype crisis response training program. The program is designed to train Army leaders – including career Army, National Guard, and Reserve officers – and their civilian counterparts in the crisis response behaviors that are required for effective performance in homeland security and natural disaster scenarios. As noted above, the training used the deliberate training technique to accelerate the development of complex cognitive behaviors (Lussier, Shadrick, & Prevou, 2003). It is a "training accelerator" that is intended to be completed prior to more costly, high-fidelity training exercises (Beaubien, et al., 2007; Shadrick, in preparation).

Development of the prototype crisis response training program followed the FLEX iterative process of knowledge elicitation, domain expert review, and rapid development. The process began with a review of documents (including executive orders, "critical incident" reports, after-action review reports, and technical manuals) provided by various state, county, and municipal government agencies. From these materials, the research team developed an initial understanding of the various types of homeland security and natural disaster crises which might occur during the coming decade, as well as the crisis response behaviors that were required to safely and effectively resolve these situations. The research team also developed an initial set of operating assumptions about how Army National Guard leaders are expected to coordinate their actions with civilian and non-governmental organizations during such crises.

Next, the research team conducted the first in a series of interviews to validate those initial assumptions and to collect additional information for use in developing the training content. The interviews included a combination of those traditional methods of knowledge elicitation already discussed (e.g., Protocol Analysis, the Delphi technique, and semi-structured interviews) as well as future oriented techniques (Landis, Fogli, & Goldberg, 1998). One purpose of these interviews was to identify examples of historically effective (and ineffective) crisis response behaviors. Another purpose was to identify events that had not yet happened, but were assessed as possible occurrences within the next few years—events for which the interviewees felt that their respective agencies were particularly ill prepared to handle.

The distributed nature of expertise in crisis situations can be seen by the makeup of the expert sample. The interviewees included 27 experts from eight different organizations, including the Army National Guard (ARNG), the ARNG Joint Operations Center, the ARNG Weapons of Mass Destruction Civil Support Team, the Department of Environmental Management, the Department of Transportation, the Department of Homeland Security, two different County Emergency Management Agencies, and the State Police. Care was taken to

sample experts from both rural and urban agencies, because such agencies differ greatly in terms of available resources and response styles.

From these interviews, the research team developed a series of high level training scenarios for potential inclusion into the training program. When designing the scenarios, special care was taken to ensure that each scenario possessed certain characteristics. First, each scenario was required to contain one assigned mission for the ARNG. Second, each scenario required extensive coordination among multiple stakeholder groups. Finally, the scenarios were dynamic, requiring individuals to critically analyze each situation to determine whether or not it was an accident, natural disaster, or homeland security incident. All of the crisis response scenarios are listed and described in Appendix A.

The CTA results were also analyzed to develop a list of behavioral themes that consistently reappeared throughout the interviews. Many of these themes are required for adaptive leaders (Shadrick & Lussier, 2004). For example, both sets of skills require maintaining a shared awareness of the commanders' intent, anticipating likely events, and positioning one's resources so that they can easily be redeployed as the situation unfolds. All of the themes are listed and described in Appendix B.

Next, scenario storyboards were developed for each of the 15 scenarios. Each storyboard included several critical pieces of information to aid in developing the multimedia-based training materials, including the following: the estimated amount of time required to present the information on screen, the specific behaviors that were targeted, the actual training content (usually presented via narration or character dialogue), and recommendations for the visual presentation via aerial photography, stock photographs, or narration. The scenarios, themes, and storyboards were then reviewed by 26 experts from the previously identified organizations. All materials were then revised based upon the expert feedback.

The researchers also developed a set of behavioral indicators that students could use to self-score their performance after responding to each training scenario. The indicators serve as benchmarks of expert performance unique to a given scenario. A set of instructor notes for each scenario was also developed. The notes provided specific probe questions to the instructor, and were designed to assess participant understanding on each of the themes. A series of key points explained the importance of each theme. The storyboards and instructor notes were assessed during a final round of expert interviews with 35 additional experts from the previously described agencies.

During each successive round of expert interviews, the researchers were able to identify and resolve critical shortcomings in the proposed training content and format. After the third round of interviews, the experts saw no more needed substantive changes, suggesting that the major issues had been resolved. This set of training materials was then used to develop the final training program. It is important to note that the expert assessments of the training materials were overwhelmingly positive, and the experts strongly supported the relevance of the materials to both the ARNG and other stakeholder groups.

Content Validation of FLEX-driven Training

Content validation procedures are used to evaluate how well important aspects of a domain are represented. If content validation procedures support content validity, they also support the method used to generate the training content including the ability of that method to tap distributed expertise and foster knowledge creation. There are various quantitative indices that can be used to assess content validity. Such indices have been frequently used to evaluate test content (Moscosco & Salgado, 2001; Wynd, Schmidt, & Schaefer, 2003), but their use in the evaluation of training content has been infrequent (Goldstein, 1980). More recent research has supported the use of these approaches in the content validation of training (Teachout, Sego, & Ford, 1997/1998; Goldstein & Ford, 2002).

Research in test content validation has also found a correlation between indices of content validity and criterion validity (Carrier, Dalessio, & Brown, 1990; Dubois & Dubois, 2000). The literature further suggests that overall content validity can be improved by using such indices to cull items (Ford & Wroten, 1984). The leveraging of such procedures is arguably even more important in tasks wherein the costs of inferior performance are high.

Conceptual Groupings and Cohen's Kappa

One aspect of the prototype crisis response training that can be metrically assessed involves the how representative the indicators are of the proposed scenarios. In other words, could individuals experienced in crisis response efforts place the indicators (i.e., those critical aspects of a situation which must be taken into account before effective decision-making and action taking can occur) into their corresponding scenarios? (Henceforth, the term "conceptual groupings" will be used to refer to the placement of indicators into particular scenarios.) Cohen's Kappa was the parametric used (Cohen, 1960). Cohen's κ is a statistic that ranges from -1 to +1. Positive values indicate levels of inter-rater agreement that is higher than expected on the basis of chance alone. Negative values, conversely, indicate the opposite (Kronk, Ogonowski, Rice, & Feldman, 2005). In general, good agreement is indicated by values of κ between .60 to .79. Values of .80 and upwards are generally considered markers of excellent agreement (Landis & Koch, 1977).

Content Validity Ratios and Content Quality

The second aspect of the training content that was quantified was component essentiality. That is, were the components (themes, scenarios, and indicators) essential for successful task performance? That question was assessed by eliciting Content Validity Ratio (CVR) judgments from each expert. Like Cohen's κ, CVRs range from -1 to +1 (Lawshe, 1975). Negative values indicate that more "not essential" than "essential" ratings were assigned to a given training component. Positive values indicate more experts judged the component to be essential than not. The CVRs of zero indicate that the experts were evenly split in their assessment of the essentiality of the component.

For confidence to be placed in these ratings, however, it was also necessary to demonstrate reliability. Therefore, the intra-class correlation coefficient (ICC) was calculated to

assess inter-rater reliability. In general, good inter-rater agreement is indicated by ICC values ranging from .45 to .75, while values greater than .75 are seen as indicating excellent agreement (Fleiss, 1981).

<center>*Method*</center>

Participants and Procedure

A sample (*n* = 6) of Army National Guard officers with significant experience in interagency crisis response efforts participated in the content validity analyses. The sample was composed of one brigadier general, two lieutenant colonels, two captains, and one lieutenant, thereby providing a range of expertise at the company, field grade, and campaign levels of command. The overarching goals of the investigation were to independently verify the conceptual groupings of the CTA method and to assess the appropriateness of the training components. Strong evidence for the content validity would also provide support for FLEX.

The participants first viewed three of the multi-media scenarios. Following the scenarios, each participant was given a list of 27 indicators randomly sampled from the scenarios just viewed. The participants were informed that there were nine indicators from each of the three scenarios, and that the indicators were listed in random order. The officers were then asked to note those scenarios in which, in their best judgment, the indicators belonged.

The participants were then shown each of the nine behavioral themes developed as a result of the CTA. Participants provided CVRs for all of the themes. Third, the participants were shown seven scenarios. After each scenario, participants were provided with a form containing the name of the scenario and a list of all of the indicators accompanying that scenario, as derived from the CTA. Each participant was then asked to provide CVRs for both the scenario and its accompanying indicators. At the end of the form, space was provided for any indicators that the experts judged to be important but missing from the current version of the training. A group discussion among all of the experts was then conducted, and when consensus was reached on an indicator suggested for inclusion, that indicator was noted.

Data Analysis

To cross-validate the CTA conceptual groupings it was necessary to compute Cohen's κ twice. The first κ assessed inter-expert agreement. If a significant level of agreement was seen, then the second κ would compare the conceptual grouping used by the majority of the experts to the CTA conceptual grouping. Significant and positive values of κ for both comparisons would constitute strong support for the CTA method.

The following decision rules were established to guide the assessment of CVR values. Significantly positive CVR values would be seen as unequivocal support for that particular item (indicator, scenario, or theme). Conversely, negative values would indicate that the item should be removed or reworded. However, non-significant positive values are more ambiguous. A non-significant (albeit positive) CVR value could indicate problems with the item, or it could be due to statistical limitations (e.g., low statistical power). Therefore, the term "qualifiedly accept" is used to refer to such items.

Results and Discussion

Conceptual Groupings and Cohen's Kappa

Results strongly supported the proposed CTA conceptual groupings. Analysis of the conceptual groupings of the experts indicated good levels of agreement (κ =.63, percent agreement = .76, p < .001). The original ratings of the judges were then examined and recoded to reflect how each indicator was categorized by the majority of the experts. (It should be noted that for each indicator, a clear cut majority decision was observed.) The resulting conceptual grouping was then compared to the CTA conceptual grouping. In this instance, the resulting κ indicated good agreement, bordering on excellent (κ = .78, percent agreement = .85, p < .001).

CVRs and Content Quality

For the themes, all of the CVRs were +1.00, indicating that they were unanimously viewed as essential. Similarly, CVR values for all seven of the scenarios reviewed were +1.00, again indicating unanimous assignment of items to the 'essential' category. When evaluating the indicators, the SMEs indicated verbally—and unequivocally—that a handful of the indicators should be removed outright, and therefore closer examination of the CVRs for those items was unnecessary. A total of 122 indicators then remained. Only one indicator was assigned a negative CVR value, indicating that it should be removed. Of the remainder, 111 were positive and 10 were neutral. Of the 111 positive CVRs, 63 were statistically significant (p < .05, one-tailed). Furthermore, the ICC for the ratings indicated a good level of inter-rater agreement (ICC= .62, p < .001). See Table 1 for CVR values assigned to expert indicators.

Table 1

CVR Frequency Values for Expert Indicators

CVRs	Frequency	Percent	Sig.
-.33	1	.8	.69
0.00	10	8.2	1.00
+.33	25	20.5	.69
+.66	23	18.9	.22
+1.00	63	51.6	.03

In addition to those indicators already present, several were suggested for inclusion in future versions of the training. One particular indicator suggested for inclusion merits further discussion. The most experienced participant suggested that crisis response personnel consider who to handle various questions from the public. The necessity of including such indicators was more apparent to the higher ranking officers. This finding underscores the necessity of including a heterogeneous sample of experts for use in knowledge creation exercises. The need for including an item may not be apparent unless experts are sampled from the appropriate domains

(e.g., crisis response efforts), or even from the appropriate levels of organizations concerned with that domain.

In general, the results support the content validity of the training. Most of the training items were assigned positive or positive and significant CVR values. As noted above, positive CVR values are associated with higher criterion validities. Therefore, there is good reason to argue for the content validity of the training. In addition, the CTA conceptual groupings overlapped significantly with the conceptual groupings of the experts used in the content validation study. In other words, the structure of the training was also supported. Finally, it should be noted that these results also support the use of FLEX. Application of FLEX to the domain of crisis response yielded training whose items and structure were supported when evaluated by an independent group of SMEs.

General Discussion

Using Content Validation to Assess FLEX

In this paper we have argued that certain domains (e.g., crisis response efforts) are not amenable to traditional CTA approaches. It is also often difficult to directly determine the criterion related validity of training content for those domains, for three overlapping reasons. First, those events are relatively infrequent. While that is fortunate from a humanitarian standpoint, it also means that the empirical bases for formulating lessons learned are sparse. Second, the focus during task operation in such cases is rightly placed upon performing the task, saving lives, and protecting property, not on determining criterion validities for test or training content. Finally, some of the tasks involved have not yet occurred. It is therefore desirable to use content validation procedures which have been shown to boost content criterion correlations in other domains (Carrier, Dalessio, & Brown, 1990; Dubois & Dubois, 2000).

Overall Assessment of FLEX

The FLEX holds considerable promise when dealing with task domains which are expected to undergo rapid change, which involve distributed expertise, or which have not yet occurred. Crisis response efforts can possess all these characteristics, and are further complicated by the need for cooperation and coordination between various civil and military agencies.

It has already been pointed out that incompatibilities in doctrine can create misunderstanding and conflict between agencies. Anecdotal evidence in support of this was provided by experts during both the development and the content validation of the training. Moreover, qualitative analysis of the data during content validation revealed that among the list of indicators were items focused upon the mutual understanding of the responsibilities and authority limits of civilian and military agencies, misunderstanding of which has in the past caused "friction" between ostensibly collaborating agencies.

Using Experts for Knowledge Elicitation and Creation

The content validation results illustrate the necessity of sampling experts with varying types of experience within a domain and utilizing experts from different domains. This raises the question: How does one identify the experts required for knowledge creation in complex and unpredictable problem spaces?

It is not always easy to identify those who are truly expert. Often experience is used as a surrogate measure of expertise. Unfortunately, expertise and experience are not synonymous (Camerer & Johnson, 1991; Rohrbaugh & Shanteau, 1999). First, expertise in one area is no guarantee of expertise in another area (Newell & Simon, 1972). That has direct implications for dynamic tasks such as crisis response or asymmetric warfare: a current expert may not be an expert on future tasks or systems. Therefore, it is no surprise that not all experts are capable of reasoning beyond "familiar" tasks and conditions. Second, knowledge gleaned from experts within a domain may be biased by over optimism. Research indicates that experts within a field tend to overestimate the importance and immediacy of breakthroughs in their own field (Tichy, 2002), such as advances in information technology and artificial intelligence.

The problem of adequately sampling expertise and balancing expert bias is acute when dealing with crisis response efforts. One approach to the problem—the one used in this research—is to develop a method which minimizes expert bias by drawing upon distributed expertise. The FLEX minimizes biases in expert reasoning by sampling experts broadly within and across domains, by iteratively reviewing expert information, and finally by utilizing consensus building procedures. In addition, although not all experts can reason beyond familiar tasks and conditions, research indicates that FLEX is capable of stimulating the production of creative, well reasoned responses in unfamiliar domains (Gossman et al., 2005). Furthermore, information derived via FLEX was able to generate training materials whose content validity was supported by experts with significant crisis response experience. Another approach to this problem might be to develop a systematic method for identifying experts. A potential starting point is the taxonomy of Bloom (1956), which highlights the ability to synthesize and apply information in familiar and unfamiliar domains.

Refinement and Extension of Content Validation Methods

Throughout this paper, we have argued that formal content validation procedures are an important means of evaluating and refining training. Although the methods used here are hardly unknown in the broader training literature (Goldstein & Ford, 2002), their use in the evaluation of Army training is infrequent. This is disappointing, because the use of content validation procedures is especially important when ensuring that the correct behaviors are trained in high stake domains (e.g., crisis response, battle command).

However, it is not enough to simply use traditional methods of content validation. Preferably, the training researcher would be able to use multiple measures to assess content validity of training. After all, the more diverse the assessment approaches used, the more likely that any existing weaknesses in training will be found. Therefore, if the application of a diverse set of content validation procedures still strongly supports the content validity of a training

program, then greater confidence can be placed in the quality of the training. Further, classical test theory states that multiple measures can reduce the likelihood of an observed relationship being either inflated due to method variance or attenuated because of the imperfect reliability of a single measure (Crawford, Henry, Ward, & Blake, 2006).

One such approach is latent variable analysis (e.g., principal components analysis, or PCA). It is possible to use latent variable approaches to analyze Likert scale ratings and recover latent (i.e, implicit) patterns (Maurer & Pierce, 1998). These patterns could then be compared to aspects of either the training program or the experts doing the rating.

As an example of how such an approach could be used to validate aspects of the training program, recall that the indicators are 'nested' in both scenarios and themes. We have demonstrated here, using Cohen's Kappa, that SMEs were able to place indicators in the correct scenarios. Suppose, instead, that we had SMEs use a 7-point Likert scale to make pairwise similarity ratings (1=not similar at all, 7=highly similar) of indicators. Would we find that those indicators which showed high and positive loadings on the same components were, in fact, indicators which belonged to the same theme?

Similarly, we might be able to assess individual organizational affiliation. During the CTA phase, it became apparent that the indicators generated by Department of Homeland Security (DHS) personnel were not identical to those generated by ARNG personnel.   If the indicators generated by the different agencies during the CTA phase are in fact truly representative of the concerns appropriate for those agencies, we should be able to demonstrate during the content validation phase that an independent sample of DHS raters "treat" the DHS generated indicators differently than those generated by ARNG personnel. To this end, we have generated a notional data set demonstrating how this could be accomplished. During the content validation phase, we could have ARNG and DHS personnel rate a random sampling of indicators as to their importance (1=not important, 7=very important). We might expect that the DHS personnel ratings which load highly and positively on the same component(s) would be ratings of items which were generated by DHS personnel during the CTA phase. A similar pattern might emerge between ARNG ratings and ARNG generated items.

The latent variable approach offers simplicity: it is a means of data reduction, and can often clarify complex relationships. Further, such approaches might also be able to serve (like the CVR) as proxy variables for direct measurement of criterion variables. Take, for example, the topic of massively multiplayer online role playing games (MMORPGs). (A related class of first person multiplayer shooters includes *America's Army*, the popular official game of the U.S. Army.)

Although there is a wealth of data that can potentially be gathered by such games, it is difficult to determine what criterion variable should be compared to game performance. However, one reasonable approach might be to select individuals with certain experiences (e.g., individuals with differing amounts or kinds of combat experience, different levels of military rank, etc.) who have not played the game. A latent variable or data reduction approach could be applied to their game data to see if any reliable patterns emerge. The goal would be to get an

idea of what the performance profile of an expert looks like. In this way, data gathered internally by the game platform itself might serve as a reasonable proxy variable for criterion performance.

Whether or not such latent factors could, in fact, be recovered from such data sets is an empirical question. Like any factor analytic technique, the ability to discover such patterns is not only due to whether or not the factors exist, but also to psychometric factors such as the composition of the participant sample, the variance explained by the factors, the number of items rated or measured, and so on. Nonetheless, the methods discussed here hold the promise of improving content validation in a variety of high stake domains relevant to both crisis response agencies and the broader domain of military training.

Conclusion

The training construction and validation approach used in this paper offers an example of how to approach training for dynamic, complex domains and future task conditions. A relevant example of such a domain is the Future Combat Systems (FCS), which are a high tech suite of sensors and other resources which promise to change the nature of warfare. Because it will be some time before the FCS systems are completed and fielded, it behooves the Army to prepare training and doctrine for the changes that FCS will bring.

While there is good reason to trust the content validity of the Red Cape training package, it must also be shown that participation in the training results in proficiency gains. That final assessment phase is underway at the time of this writing: training sessions with the training package are being conducted with audiences drawn from diverse relevant agencies. We hypothesize that deliberate training in applying the themes of crisis response to situations results in the increased ability to 'size up' situations—that is, to extract information needed for appropriate action taking and decision-making.

The qualitative participant feedback gathered so far is promising. Many of the comments elicited from participants focused upon how realistic the scenarios were (e.g., several participants stated that "These scenarios could easily happen"). Perhaps the most beneficial aspect of the training, however, is the interagency discussions which take place after a scenario has been assessed at the individual level. Many participants highly praised this aspect of the training. For example, participants stated that "The training opened my eyes to areas in which I have little experience" and "The discussion…encouraged all to think outside the box."

Other data elicited from the participants indicates that increasing experience with the training resulted in increased appreciation of the training. The comparison of Likert scale responses elicited near the beginning and midpoints of the training suggest that the training was increasingly seen as effective and well designed. Interestingly, there is some (albeit far from conclusive) evidence that the experts may have been prone to overconfidence before the training began. Later Likert scale ratings show that the scenarios were seen as increasingly challenging, perhaps indicating an increased awareness of the complexities involved in large scale interagency efforts. In particular, a scenario focused upon the aftermath of a wide scale earthquake was met by initial stunned silence on the part of the participants. One participant from the Department of Homeland Security stated (paraphrased) "Too often, persons from DHS

and other non-military agencies assume that the military will have all personnel and equipment available. These interactions are useful as a reality check." In sum, the Red Cape training package appears to live up to its promise—the delivery of low cost, effective, deliberate theme-based training in the domain of crisis response.

References

Aguilar, F. J. (1967). *Scanning the business environment*. New York, NY: Macmillan Co.

Beaubien, J., Paley, M., Badugu, S., Shadrick, S. B., Ennis, C., & Jacklin, S. (2007). *Red Cape: Crisis Action Planning and Execution*. (Research Product 2007-01). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Bloodgood, J. M., & Salisbury, W. D. (2001). Understanding the influence of organizational change strategies on information technology and knowledge management strategies. *Decision Support Systems, 31,* 55-69.

Bloom, B. S. (Ed.) (1956). *Taxonomy of educational objectives: The classification of educational goals*. New York: Longmans, Green.

Camerer, C. F., & Johnson, E. J. (1991). The process-performance paradox in expert judgment: How can the experts know so much and predict so badly? In K. A. Ericcson & J. Smith (Eds.), *Towards a general theory of expertise: Prospects and limits* (pp. 195-217). Cambridge, U. K.: Cambridge University Press.

Carrier, M. R., Dalessio, A. T., & Brown, S. H. (1990). Correspondence between estimates of content and criterion-related validity values. *Personnel Psychology, 43* (1), 85-100.

Charness, N., Tuffiash, M., & Krampe, R. (2005). The role of deliberate practice in chess expertise. *Applied Cognitive Psychology, 19* (2), 151-165.

Chervinskaya, K. R., & Wasserman, E. L. (2000). Some methodological aspects of tacit knowledge elicitation. *Journal of Experimental & Theoretical Artificial Intelligence*, *12*, 43-55.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37-46.

Crawford, J. R., Henry, J. D., Ward, A. L., & Blake, J. (2006). The Prospective and Retrospective Memory Questionnaire (PRMQ): Latent structure, normative data and discrepancy analysis for proxy-ratings. *British Journal of Clinical Psychology, 45* (1), 83-104.

Dalkey, N. C. (1969). *The Delphi method: An experimental study of group opinion*. Santa Monica, CA: Rand Corporation.

Dixon, R. G. (2006). *Systems thinking for integrated operations: Introducing a systematic approach to operational art for disaster relief*. Fort Leavenworth, KS: School of Advanced Military Studies (DTIC No. ADA448434)

Dubois, D. A., & Dubois, C. L. Z. (2000). An alternate method for content-oriented test construction: An empirical evaluation. *Journal of Business and Psychology, 15* (2), 197-213.

Dubois, D. A., Shalin, V. Levi, & Borman, W. C. (1995). *A cognitively-oriented approach to task analysis and test development*. Minneapolis, MN: Personnel Decisions Research Institute, Inc. (DTIC No. ADA308838).

Ericsson, K. A., Krampe, R. T., & Tesch- Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100* (3), 363-406.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87*, 215-253.

Executive Office of the President: National Science and Technology Council. (2006). *Combating terrorism: Research priorities in the social, behavioral and economic sciences*. Washington, DC: Author.

Fallows, J. (2006). Declaring victory. *The Atlantic*, *298* (2)*,* 60-75.

Farmer, L. C. & Williams, G. R. (2005, October). *The rigorous application of deliberate practice methods in skills courses*. Paper presented at the UCLA/IALS Sixth International Clinical Conference on Enriching Clinical Education, Lake Arrowhead, California.

Fleiss, J. L. (1981). *Statistical methods for rates and proportions*. New York: Wiley.

Ford, J. K., & Wroten, S. P. (1984). Introducing new methods for conducting training evaluation and for linking training evaluations to program redesign. *Personnel Psychology, 37*, 651-665.

Glenn, R. W., Jacobs, J., Nichiporuk, B., Paul, C., Raymond, B., Steeb, R., & Thie, H. J. (2006*). Preparing for the proven inevitable: An urban operations training strategy for America's joint force*. Santa Monica, CA: RAND Corporation.

Goldstein, I. L. (1980). Training in work organizations. *Annual Review of Psychology, 31*, 229-272.

Goldstein, I. L., & Ford, J. K. (2002). *Training in organizations: Needs assessment, development, and evaluation* (4th Ed.). Belmont, CA: Wadsworth.

Gossman, J. R., Mauzy, R. P., Heiden, C. G., Campbell, C. H., Flynn, M. R., Lussier, J. W., & Shadrick, S. B. (2005). *Flexible methods for Future Force concept development*. (Research Report 1839). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Knapp, D. J., Burnfield, J. L., Sager, C. E., Waugh, G. W., Campbell, J. P., Reeve, C. L., Campbell, R. C., White, L. A., & Heffner, T. S. (2002). *Development of Predictor and Criterion Measures for the NCO21 Research Program.* (Research Report 1128). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Kronk, R. A., Ogonowski, J. A., Rice, C. N., & Feldman, H. M. (2005). Reliability in assigning ICF codes to children with special health care needs using a developmentally structured interview. *Disability and Rehabilitation, 27* (17), 977-983.

Landis, K. R., & Koch, G. G. (1977). The measure of observer agreement for categorical data. *Biometrics, 33*, 59-174.

Landis, R. S., Fogli, L., & Goldberg, E. (1998). Future-oriented job analysis: A description of the process and its organizational implications. *International Journal of Selection and Assessment 6* (3), 192-197.

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology, 28*, 563-575.

Lussier, J. W., Shadrick, S. B., & Prevou, M. I. (2003). *Think Like a Commander prototype: Instructor's guide to adaptive thinking.* (Research Product 2003-02). Alexandria, VA: U. S. Army Research Institute for the Behavioral and Social Sciences.

Meredith, J. R., & Mantel, S. J. (2004) Technological Forecasting (Appendix). In Mantel, S. J., Meredith, J. R., Shafer, S. M., & Sutton, M. M. (2004). *Core concepts: Project Management in Practice* (2nd Ed.). Hoboken, NJ: John Wiley & Sons.

Maurer, T. J., & Pierce, H. R. (1998). A comparison of Likert scale and traditional measures of self-efficacy. *Journal of Applied Psychology, 83* (2), 324-329.

Meyer, M. A., & Booker, J. M. (1991). *Eliciting and analyzing expert judgment: Practical Guide.* San Diego, CA: Academic Press.

Moscosco, S., & Salgado, J. F. (2001). Psychometric properties of a structured behavioral interview to hire private security personnel. *Journal of Business and Psychology, 16* (1), 51-59.

Newell, A., & Simon, H. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice-Hall.

Nonaka, I., & Takeuchi, H. (1995). *The knowledge creating company.* New York: Oxford University Press.

Robinson, J. (1982). Energy backcasting: A proposed method of policy analysis. *Energy Policy 10* (4), 337-344.

Rohrbaugh, C. C., & Shanteau, J. (1999). Context, process, and experience: Research on applied judgment and decision making. In F. T. Durso, R. S. Nickerson, & M. T. H. Chi (Eds.), *Handbook of applied cognition* (pp. 115-139). New York: John Wiley.

Schraagen, J., Chipman, S., & Shalin, V. (Eds.) (2000). *Cognitive task analysis*. Mahwah, NJ: Lawrence Erlbaum Associates.

Shadrick, S. B. (in preparation). *Red Cape: Crisis response training for National Guard and interagency partners*. Arlington, VA: U. S. Army Research Institute for the Behavioral and Social Sciences.

Shadrick, S. B., & Lussier, J. W. (2004). *Assessment of the Think Like a Commander training program*. (Research Report 1824). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Shadrick, S. B., Lussier, J. W., & Hinkle, R. (2005). *Concept development for future domains: A new method of knowledge elicitation*. (Technical Report 1167). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Shadrick, S. B., Lussier, J. W., & Fultz, C. (2007). *Accelerating the development of adaptive performance: Validating the Think Like a Commander training.*. Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Teachout, M. S., Sego, D. J., & Ford, J. K. (1997/1998). An integrated approach to summative evaluation for facilitating training course improvement. *Training Research Journal, 3*, 169-184.

Tichy, G. (2002). *Over-optimism among experts in assessment and foresight*. (Manuscript ITA-02-05). Vienna, Austria: Institute of Technology Assessment. Retrieved October 2, 2006, from http://www.oeaw.ac.at/ita/pdf/ita_02_05.pdf

Wei, J., & Salvendy, G. (2004). The cognitive task analysis methods for job and task design: Review and reappraisal. *Behaviour & Information Technology, 23* (4), 273-299.

Woods, D., & Dekker, S. (2000). Anticipating the effects of technological change: A new era of dynamics for human factors. *Theoretical Issues in Ergonomics Science*, 2000, 1(3), 272-282.

Wright, G., & Ayton, P. (1987). Eliciting and modeling expert knowledge. *Decision Support Systems, 3,* 13-26.

Wynd, C. A., Schmidt, B., & Schaefer, M. A. (2003). Two quantitative approaches for estimating content validity. *The Western Journal of Nursing Research, 25* (5), 508-518.

Appendix A

Crisis Management Scenarios

Scenario 1: Power Grid Shutdown
Wabash Valley Power Company supplies electricity to 27 distribution cooperatives throughout
northern and central Indiana. The company recently experienced a shutdown at one of those
cooperatives. A power grid failure resulted, thereby shutting down electrical power to two-thirds
of the state. The shutdown occurs during a severe major winter storm, leaving many residents, at
risk of hypothermia.

Scenario 2: Industrial Plant Explosion
A major explosion occurred this morning at a chemical plant in Gary, IN. The plant is located
adjacent to Interstate 80/90, adjacent to the Grand Calumet River, one quarter mile south of Lake
Michigan, one mile north of the Methodist Hospital Northlake Campus, and 13 miles east of the
Illinois border. Electrical power, telephone and cable television interruptions are being reported
throughout the Gary metropolitan area.

Scenario 3: Capital Punishment of High Profile Prisoner
In May 2000, a federal jury in Kansas City, MO, recommended a death sentence for Manuel
Riley of Houston, TX, for his role as triggerman in the murder of a drug dealer. Investigations
conducted during the trial established a link between Riley and the terrorist group ARF (Anarchy
Revolutionary for Freedom). Riley's pending execution provides an opportunity for the ARF to
orchestrate a series of terrorist attacks under the guise of legitimate protest.

Scenario 4: Dirty Bomb
An automobile packed with explosives exploded in downtown Indianapolis. The explosion
killed a police officer and wounded 12 citizens. As the initial explosion aftershock subsided, a
"talc" like powder was seen covering the blast zone. Area citizens rushed to the scene and
attempted to render assistance. Local merchants opened their establishments for any relief effort.
Emergency medical crews were transporting the injured to the local hospitals. A terrorist group
later claimed responsibility for the attack.

Scenario 5: Vehicle with Hazardous Material (HAZMAT)
The Indiana State Police receive a 911 call from a motorist with a cell phone.   A commercial
truck with a tandem semi-trailer has overturned in the eastbound lane of I-64 between mileposts
32 and 33. The rear tanker trailer is on its side, blocking both lanes at the bridge over Pigeon
Creek and leaking heavily from the top hatch. The tractor and front tanker trailer left the
highway and went down the embankment, coming to a rest on its side in Pigeon Creek. The
caller can see a hazardous cargo placard on the overturned trailer, but does not want to venture
closer to the vehicle to determine what it was carrying.

Scenario 6: Severe Earthquake along New Madrid Fault
A magnitude 7.5 earthquake has occurred along the New Madrid fault, impacting a multi-state
region that includes Arkansas, Illinois, Indiana, Kentucky, Mississippi, Missouri, and Tennessee.
Reports from Evansville, Clarksville, and Bloomington indicate catastrophic damage in all three

cities. Buildings have collapsed and debris is blocking city streets. Numerous fires are reported burning in each city, and water main breaks have significantly hindered firefighting operations. Casualty and death figures are sketchy but growing rapidly. Hospitals are functioning on generator power and are overwhelmed with victims. Electricity and water are out in major portions of each city.

Scenario 7: Sports Riot in a University Town
In late August, Indiana University is playing host to Notre Dame in an early season football game. Two weeks prior, ESPN contacted the university and announced that they were going to make this a nationally televised football game. An unrelated 3-day outdoor rock concert had been previously scheduled outside the city limits, with performances scheduled for 15 hours per day on Thursday, Friday, and Saturday. Small conflicts between the football and music fans, eventually reach a boiling point.

Scenario 8: Storm of the Century
On Thanksgiving morning, the state of Indiana was hit with a major storm event. The lake area received 24 inches of snow, in an 18 hour period. The middle section of the state was hit by the same storm system dumping 3.5 inches of ice on the area. Due to the previous weeks' heavy rain, the Ohio River area was showing signs of potential flooding. The flooding conditions are only exacerbated by the current storm. The Army Corps of Engineers predicts for the river to crest in 18 hours around the counties of Clark, Harrison and Spencer. The storm results in numerous problems that hamper health care, first response, and public works organizations.

Scenario 9: Prison Riot with Helicopter Crash
A fight in the dining facility at the Indiana State Prison in Michigan City has escalated into a full scale riot. Several prison guards have been overpowered and multiple wings of the prison have been unlocked. Prisoners are escaping into the halls, administrative areas, and the prison yard. Initial reports indicate that three guards have been severely beaten and at least five prisoners killed. As a negations team is flown in via helicopter. However, their tail rotor fails during the approach and crashes inside the perimeter.

Scenario 10: Nuclear Bomb in Shipping Container
Intelligence reports indicate that Al Qaeda has obtained a five kiloton nuclear weapon and the associated activation codes. That size bomb would decimate a major city the size of New York or Chicago. During the past three weeks, an increase in the volume of cell phone and e-mail intercepts has led the Department of Homeland Security to believe that plans are being made to deploy the bomb. Six days ago, all communications traffic suddenly ceased. The Department of Homeland Security believes that Al Qaeda has gone into "radio silence" and that the bomb may be in transit. The Burns Harbor is identified as a likely point of entry for the weapon.

Scenario 11: Airplane Crash in Restricted Area
Forty five minutes ago, a 12-passenger corporate jet crashed just short of the U.S. Army Newport Chemical Weapons Depot. The jet's flight plan indicated that it was scheduled to fly from Kansas City, MO, to Cincinnati, OH. Twelve minutes prior to the crash, the jet veered 30 degrees off course and began descending from 30,000 feet. No emergency was declared and the

FAA was unable to contact the pilot over any radio frequency. Because the crash crossed the Depot fence line, multiple jurisdictions are involved.

Scenario 12: Animal Borne Disease at Stockyard
A worker at a cattle and livestock show has notified the Indiana State authorities that a breeder cow is displaying signs of mad cow disease. Because the virus spreads very quickly, it is imperative to track down everyone who attended the sale, so that their clothes can be decontaminated and their herds tested. The virus is traced back to a local cattle farmer who refuses to allow inspectors onto his farm.

Scenario 13: Industrial Plant Fire near INANG Headquarters
At approximately 8:15 a.m. (local time) an explosion occurred at a major industrial park located off I-70 just west of downtown Indianapolis (Holt Road exit). Flying debris from the explosion ruptured a Paranitrotoluene (PNT) storage tank causing a fire. The fire has released a plume of toxic gas that is headed straight for the Indiana National Guard Joint Forces headquarters.

Scenario 14: Rail Yard Explosion
The Indiana Army National Guard is simultaneously deploying two battalions to the Middle East. As the units progress through their qualification and pre-deployment procedures, both are required to load their vehicles onto rail cars for transport to the port in Charleston, SC. An explosion occurs during the loading procedure, killing several Guard Soldiers and damaging multiple vehicles.

Scenario 15: INANG Arrives in Theater
After 21 days of unit training and preparation, the 4th of the 42d Armor Battalion of the Indiana National Guard arrives in the Middle East for a 12-month rotation. Shortly thereafter, the battalion moves by tactical convoy to the town of Khan al Sur near the northwest border of Iraq and Syria. Their mission is to prevent insurgents from crossing over the Syrian border and assist a newly formed Iraqi infantry battalion in becoming combat proficient. Multiple problems immediately ensue.

Appendix B

Crisis Management Themes

Maintain Focus on Mission Priorities
Effective managers recognize that their primary mission is to protect human life and property, with safety being paramount. Plans and actions should focus on preserving life and property while also addressing the incident and its underlying causes. Effective managers may choose to delay, deflect, or delegate activities that address secondary or tertiary objectives.

Keep the Chain of Command Flexible
Effective managers recognize that the chain of command does not remain fixed throughout the crisis. Management roles often evolve as the crisis unfolds and additional agencies become involved. Effective managers recognize and transition between manager and subordinate (or participant and observer) as the crisis unfolds.

See the Big Picture
Effective managers remain aware of what is happening around them, with particular attention of the impact on higher, adjacent, lower, and supported organizations. Effective managers attempt to analyze all factors inside and outside the organization's direct control that could affect its activities during the time line covered by the plan, to include geographical scope, population involvement, incident duration, and lead time required for particular actions.

Plan for and Recognize Decision (Trigger) Points
Effective managers identify key and measurable indicators that require immediate action. They develop plans of action for these trigger points to avoid premature commitment of key assets and failing to act in a timely manner.

Reprioritize as Necessary
Effective managers continually reassess the scope and priority of mission requirements as the crisis unfolds. They continually validate new information as it becomes available, and adjust their mission plans as the crisis unfolds and approaches resolution. Effective managers are able to determine "time-critical" versus "resource-critical" requirements and adjust the plan accordingly.

Use All Available Assets
Effective managers remain aware of what assets are available to them, what their limitations are, and what lead times are required. They listen to the input of subject matter experts and are prepared to defer judgment to them, as necessary. Effective managers use persuasion and cooperation in situations where directive authority is not appropriate.

Think in Shades of Gray, not Black and White
Effective managers recognize that they will be working with imperfect and incomplete information from a wide range of organizations and individuals. They consider the source of the data and allow for variations when applying that data.

Model a Dynamic Situation

Effective managers recognize that the cause of the incident may be an accident or act of nature, or it may be either criminal or terrorist activity. They avoid the temptation to simplify the situation by treating the cause of the crisis as static or simply reactive.

Understand the Public Need

Effective managers recognize that the public will experience a wide range of emotions, from proactive attempts at involvement to a passive acceptance of the situation. Certain reactions could lead to irrational behavior from the public. The effective manager takes the time to assess these reactions, explain (when appropriate) the reasoning behind certain actions and attempts to allay any public fear.